

14

Jc848 U.S. PTO

C-862 U.S. PTO
09/625381



07/25/00

(For new non-provisional applications under 37 CFR § 1.53(b))

Atty. Dkt. No:	5181-55200
Inventor(s):	Whay S. Lee
Title:	System and Method for Implementing a Routing Scheme Using Intention Packets in a Computer Network

CERTIFICATE OF EXPRESS MAIL
UNDER 37 C.F.R. §1.10

"Express Mail" mailing label number: EL588940106US
DATE OF DEPOSIT: 7/25/00

I hereby certify that this paper or fee is being deposited with the United States Postal Service "Express Mail Post Office to Addressee" service under 37 C.F.R. §1.10 on the date indicated above and is addressed to:

Assistant Commissioner for Patents
Box Patent Application
Washington, DC 20231


Debra J. Tix

1. ☒ Filing Fee

Total Claims	33	-20=	13	x \$18.00=	\$234.00
Independent Claims	3	-3 =	0	x \$78.00=	\$0.00
Multiple Dependent Claims				Fee:	\$0.00
Basic Fee:					\$690.00
Assignment Fee					\$40.00
Sub-Total:					\$964.00
Reduction by 50% for Small Entity:					\$0.00
Total:					\$964.00

The Commissioner is hereby authorized to charge any other fees which may be required or credit any overpayment to Conley, Rose, & Tayon, P.C., Deposit Account No. 03-2769/5181-55200/RCK.

One duplicate copy of this form is enclosed.

2. ☒ Specification
19 page(s) of specification; 7 page(s) of claims, 1 page(s) of abstract
3. ☒ Drawings
 Formal Figure(s) 1-7 on 7 sheet(s)
4. ☒ Oath or Declaration
☒ Newly executed
☐ Copy from a prior application (see 37 C.F.R. § 1.63(d))
 Deletion of Inventor(s) (in continuation or divisional applications)
☐ Delete the following inventor(s) named in the prior non-provisional application:

☐ The inventor(s) to be deleted are set forth on a signed sheet attached to this application.

5. ☐ The entire disclosure of the prior application referred to above is considered to be part of the accompanying application and is hereby incorporated by reference herein.
6. ☐ Microfiche Computer Program (Appendix)
7. ☐ Nucleotide and/or Amino Acid Sequence Submission (if applicable, all necessary)
- ☐ Computer Readable copy
- ☐ Paper Copy (identical to computer copy)
- ☐ Statement verifying identity of above copies
8. ☒ Assignment Papers
9. Power of Attorney
- ☒ Is attached.
- ☐ The power of attorney appears in the original papers of the prior application.
- ☐ Since the power does not appear in the original papers, a copy of the power in the prior application is enclosed.
10. ☐ Information Disclosure Statement (IDS)
- ☐ Copies of IDS Citations
11. Amendments
- ☐ A preliminary amendment is enclosed.
- ☐ Cancel in this application claim(s) _____ before calculating the filing fee. At least one independent claim is retained for filing purposes.
- ☐ Amend the specification by inserting before the first line the sentence: _____.
12. ☒ Return Receipt Postcard
13. Small Entity Status
- ☐ A small entity statement is enclosed.
- ☐ A small entity statement was filed in the prior non-provisional application and such status is still proper and desired.
- ☐ Is no longer claimed.
14. ☐ Priority of foreign application number _____, filed on _____ in _____ is claimed under 35 U.S.C. §§ 119(a)-(d)
15. ☐ Petition under 37 C.F.R. § 136 for Extension of Time
16. ☐ Other: _____

Address all future correspondence to:

Robert C. Kowert
Conley, Rose, & Tayon, P.C.
P.O. Box 398
Austin, Texas 78767
Phone: (512) 476-1400 Fax: (512) 703-1250

Signature

Name

Registration No.

Date

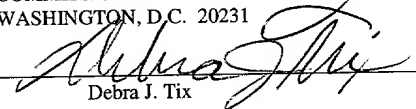

Robert C. Kowert

39,255

PATENT
5181-55200
P4639

"EXPRESS MAIL" MAILING LABEL NUMBER
EL588940106 US

DATE OF DEPOSIT JULY 25, 2000
I HEREBY CERTIFY THAT THIS PAPER OR
FEE IS BEING DEPOSITED WITH THE
UNITED STATES POSTAL SERVICE
"EXPRESS MAIL POST OFFICE TO
ADDRESSEE" SERVICE UNDER 37 C.F.R. §
1.10 ON THE DATE INDICATED ABOVE AND
IS ADDRESSED TO THE ASSISTANT
COMMISSIONER FOR PATENTS,
WASHINGTON, D.C. 20231


Debra J. Tix

System and Method for Implementing a Routing Scheme
Using Intention Packets in a Computer Network

By:

Whay S. Lee

Atty. Dkt. No.: 5181-55200/P4639

Robert C. Kowert
Conley, Rose & Tayon, P.C.
P.O. Box 398
Austin, TX 78767-0398
Ph. (512) 476-1400

005220" T8E52960

BACKGROUND OF THE INVENTION

1. Cross-Reference to Related Application

5 This application relates to co-pending U.S. patent application, serial no. _____ (Attorney Reference No: 5181-55000), filed on the same day as the present application and entitled "System and Method for Implementing a Routing Scheme in a Computer Network Using Intention Packets When Fault Conditions Are Detected" by Whay S. Lee, which is incorporated herein by reference in its entirety, and to co-pending
10 U.S. patent application, serial no. _____ (Attorney Reference No: 5181-63400), filed on the same day as the present application and entitled "Routing Mechanism Using Intention Packets in a Hierarchy of Networks" by Whay S. Lee, which is incorporated herein by reference in its entirety.

15 2. Field of the Invention

 This invention relates generally to the field of computer systems and, more particularly to communications protocols within computer systems and/or networks, and communication routing or switching within computer systems and/or networks.

20

3. Description of the Related Art

 Devices in a communications network may communicate with one another by sending packets to each other. These packets may correspond to an operation that one
25 device seeks to perform with the assistance of another device. Packets may be transferred between devices using a network which may include a fabric. A network generally may comprise a number of devices that are configured to convey packets between any number of devices.

At times, however, adverse transmission conditions may occur in the network when one device is sending a packet to another device on the network. For example, the network may become overly congested with packets traveling between devices. Another adverse condition that may occur in networks is corruption of a packet (e.g. an ECC error, parity error, checksum error, etc.). A fault condition (e.g. failure in the fabric) that prevents a packet from proceeding along its proposed route is another adverse condition. One adverse condition resolution scheme in packet routing protocols involves dropping packets in response to an adverse transmission condition within one or more devices (e.g. a switch, router, endpoint, node, etc.) in the network. In the case of congestion, a device within the network may drop one or more packets randomly or selectively to instantaneously reduce the packet load. By dropping one or more packets, the device can free up sufficient resources to service the remaining packets. Corrupted packets may also be dropped. For a fault condition, all packets reaching a fault point may have to be dropped and resent using a different route.

When a packet is dropped, the device that originated the packet needs to detect that it has been dropped so that it can be resent. Detection of dropped packets may be accomplished by a time-out mechanism at a receiving device when it fails to receive an expected packet, by a time-out mechanism at a sending device where the sending device is waiting for a signal from a receiving device that a packet has been received, or by sending a negative acknowledgement (NACK) from the device that dropped the packet to the sending device.

Each of these detection methods has its problems, however. Where a time-out mechanism is used, a device needs to ensure that it does not resend a packet until it is certain that the original packet has been dropped. The device must generally wait a period of time to ensure that the packet has been dropped and this waiting period results

in a performance penalty. Where a NACK to the sending device is used to signal a dropped packet, a dropped packet is unknown to the receiving device, no matter how far through the network the dropped packet traveled before being dropped. The NACK propagates back only to the sender and the network resources consumed by the dropped packet are wasted. An apparatus and method of handling adverse transmission conditions that minimizes the drawbacks of the above methods is needed.

SUMMARY

The problems outlined above may be solved by the use of the apparatus and method described herein. Generally speaking, an apparatus and method for implementing a routing scheme using intention packets is contemplated. At times, a switching device within a network may detect an adverse transmission condition such as a fault, corruption, or congestion. When a switching device detects an adverse transmission condition, it may drop one or more packets to alleviate the adverse condition.

For example, in the case of congestion, the switching devices may support a particular amount of routing resource (e.g. bandwidth) in and out of each of their ports. When a packet or a header portion of a packet arrives at a switching device, the switching device may determine what port the packet will need and the amount of resources required by the packet on that port. If the required resources are available for the packet on the port, then the switching device may route the packet to a next device. If the required resources are not available for the packet on the port, then the switching device may drop at least a portion of the packet. A switching device may be configured to remember the amount of resources that it has committed to other traffic and may drop a subsequent packet when it is unable to guarantee the resources requested by the subsequent packet.

When a packet is dropped due to an adverse transmission condition, the switching device that dropped it may create a second packet to be forward to the destination device. The second packet may be an intention packet that corresponds to the dropped packet. The switching device may then route or forward the intention packet to the destination device specified by the dropped packet. The intention packet may be much smaller than the dropped packet and may include a header portion or part of a header portion of the dropped packet, as discussed below. Accordingly, it may require substantially less

bandwidth than the dropped packet so that a switching device may be able to guarantee its delivery to the next device if congestion was encountered. For fault conditions, the device may send the intention packet to the destination using a different route in order to avoid the fault condition.

5

A device may react to receiving an intention packet based on the information in the intention packet as well as its own characteristics. The information in the intention packet may indicate whether the dropped packet was a request or a response to a request. If the dropped packet was a request, then the device that receives the intention packet, i.e. the receiving device, may send a signal to the requesting device that its request has been dropped. The receiving device may also perform an operation that corresponds to the request in anticipation of receiving the resent packet. If the dropped packet was a response to a request, then the requesting device will receive the intention packet. The requesting device may reschedule the request in response to receiving the intention packet and may convey a subsequent packet that includes the request to the receiving device. The requesting device may be configured to optimize the resending of the request by rescheduling the request for a more appropriate time or by merging the request with a subsequent request.

20

BRIEF DESCRIPTION OF THE DRAWINGS

Fig. 1 is a block diagram illustrating one embodiment of devices connected through a network.

5

Fig. 2a is a block diagram illustrating one embodiment of a packet.

Fig. 2b is a block diagram illustrating one embodiment of a packet.

10 Fig. 3 is a flow chart illustrating one embodiment of a method for implementing an intention packet routing scheme in a switching device.

Fig. 4 is a flow chart illustrating one embodiment of a method for handling packets at a first exemplary device.

15

Fig. 5 is a flow chart illustrating one embodiment of a method for handling packets at a second exemplary device.

20 Fig. 6 is a block diagram of one embodiment of a system including one or more networks.

Fig. 7 is a block diagram of one embodiment of a plurality of nodes connected in a network.

25 While the invention is described herein by way of example for several embodiments and illustrative drawings, those skilled in the art will recognize that the invention is not limited to the embodiments or drawings described. It should be understood, that the drawings and detailed description thereto are not intended to limit the

	1970	1971	1972	1973	1974	1975	1976	1977	1978	1979	1980	1981	1982	1983	1984	1985	1986	1987	1988	1989	1990	1991	1992	1993	1994	1995	1996	1997	1998	1999	2000	2001	2002	2003	2004	2005	2006	2007	2008	2009	2010	2011	2012	2013	2014	2015	2016	2017	2018	2019	2020	2021	2022	2023	2024	2025	2026	2027	2028	2029	2030	2031	2032	2033	2034	2035	2036	2037	2038	2039	2040	2041	2042	2043	2044	2045	2046	2047	2048	2049	2050	2051	2052	2053	2054	2055	2056	2057	2058	2059	2060	2061	2062	2063	2064	2065	2066	2067	2068	2069	2070	2071	2072	2073	2074	2075	2076	2077	2078	2079	2080	2081	2082	2083	2084	2085	2086	2087	2088	2089	2090	2091	2092	2093	2094	2095	2096	2097	2098	2099	2100	2101	2102	2103	2104	2105	2106	2107	2108	2109	2110	2111	2112	2113	2114	2115	2116	2117	2118	2119	2120	2121	2122	2123	2124	2125	2126	2127	2128	2129	2130	2131	2132	2133	2134	2135	2136	2137	2138	2139	2140	2141	2142	2143	2144	2145	2146	2147	2148	2149	2150	2151	2152	2153	2154	2155	2156	2157	2158	2159	2160	2161	2162	2163	2164	2165	2166	2167	2168	2169	2170	2171	2172	2173	2174	2175	2176	2177	2178	2179	2180	2181	2182	2183	2184	2185	2186	2187	2188	2189	2190	2191	2192	2193	2194	2195	2196	2197	2198	2199	2200	2201	2202	2203	2204	2205	2206	2207	2208	2209	2210	2211	2212	2213	2214	2215	2216	2217	2218	2219	2220	2221	2222	2223	2224	2225	2226	2227	2228	2229	2230	2231	2232	2233	2234	2235	2236	2237	2238	2239	2240	2241	2242	2243	2244	2245	2246	2247	2248	2249	2250	2251	2252	2253	2254	2255	2256	2257	2258	2259	2260	2261	2262	2263	2264	2265	2266	2267	2268	2269	2270	2271	2272	2273	2274	2275	2276	2277	2278	2279	2280	2281	2282	2283	2284	2285	2286	2287	2288	2289	2290	2291	2292	2293	2294	2295	2296	2297	2298	2299	2300	2301	2302	2303	2304	2305	2306	2307	2308	2309	2310	2311	2312	2313	2314	2315	2316	2317	2318	2319	2320	2321	2322	2323	2324	2325	2326	2327	2328	2329	2330	2331	2332	2333	2334	2335	2336	2337	2338	2339	2340	2341	2342	2343	2344	2345	2346	2347	2348	2349	2350	2351	2352	2353	2354	2355	2356	2357	2358	2359	2360	2361	2362	2363	2364	2365	2366	2367	2368	2369	2370	2371	2372	2373	2374	2375	2376	2377	2378	2379	2380	2381	2382	2383	2384	2385	2386	2387	2388	2389	2390	2391	2392	2393	2394	2395	2396	2397	2398	2399	2400	2401	2402	2403	2404	2405	2406	2407	2408	2409	2410	2411	2412	2413	2414	2415	2416	2417	2418	2419	2420	2421	2422	2
--	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	---

DETAILED DESCRIPTION OF AN EMBODIMENT

Turning now to Fig. 1, a block diagram illustrating one embodiment of devices connected through a network. Fig. 1 depicts device 100 coupled to device 120 through
5 network 110. Network 110 includes switching devices 112a through 112(n). It can be noted that any number, n, of switching devices may be included in network 110 and that the switching devices may be connected in any manner. It can be noted that device 100 and device 120 can comprise any type of devices configured to communicate with other devices. In one particular embodiment, device 100 may comprise a computer system host
10 interface to a storage system and device 120 may comprise a disk drive. Alternatively, devices 100 and 120 may be any combination of networked devices, such as client computers, servers, storage devices, etc. Network 110 may be a network on any scale, such as from the Internet, to a LAN or a network inside a single system interconnecting multiple disk drives or other I/O devices. In an Internet embodiment, device 100 may be
15 a requestor of a web page and device 120 may be a web server.

Devices 100 and 120 may be configured to communicate through network 110. Device 100 may convey a request to switching device 112a. The request may be routed or transmitted through network 110 until it is received at switching device 112(n) and
20 ultimately at device 120. Device 120, depending on the request, may convey a response to the request to switching device 112(n). The response may be routed through network 110 until it is received at switching device 112a and ultimately at device 120. In one embodiment, switching devices 112 are simple switches that include at least one input port, one output port, and operate to pass data from an input port to an output port. In
25 other embodiments, switching devices 112 may be more intelligent devices such as packet routers or computer systems.

Generally speaking, a request may be a signal from a first device to a second device for the second device to perform an operation or operations. For example, a request may signal a device to read data from or write data to a particular location. A response may be a signal from the second device to the first device that corresponds to the request. A response, for example, may indicate whether or not the operation signaled by the request was performed successfully or may convey data corresponding to the request. In one embodiment, requests and responses may be conveyed using packets defined by a communications protocol. Devices 100 and 120 may be configured to convert their respective requests and responses into packets prior to conveying the packets to network 110.

The switching devices in network 110 may implement a transmission scheme using intention packets. At times, a switching device 112 within a network 110 may detect an adverse transmission condition such as a fault, corruption, or congestion. When a switching device 112 detects an adverse transmission condition, it may drop one or more packets to alleviate the adverse condition.

Switching devices 112a through 112(n) may each include one or more ports. The switching devices may support a particular bandwidth (and/or other routing resources) in and out of each of their ports. When a packet or a header portion of a packet arrives at a switching device, the switching device may determine what port the packet will need and the amount of routing resources required by the packet on that port. If the required routing resources are available for the packet on the port, e.g. the required resources are less than or equal to the available resources, then the switching device may route the packet to a next device. If the required resources are not available for the packet on the port, then the switching device may drop at least a portion of the packet. A switching device may be configured to remember the amount of routing resources that it has committed to other traffic, i.e. other packets, and may drop a subsequent packet when it is

unable to guarantee the resources requested by the subsequent packet. A switching device may also drop corrupted packets, e.g. packets with a ECC, parity, or checksum error. A switching device may also drop packets for other adverse conditions such as a fault condition that prevents a packet from being routed along its proposed route.

5

When a packet is dropped, the switching device that dropped it may create an intention packet that corresponds to the dropped packet. The switching device may then route the intention packet to the destination device specified by the dropped packet. The intention packet may be much smaller than the dropped packet and may include a header portion or part of a header portion of the dropped packet, as discussed below. Accordingly, it may require substantially less routing resources than the dropped packet so that a switching device may be able to guarantee its delivery to the next device. In some cases, such as fault conditions, alternate paths may have to be tried for the intention packet to reach its destination.

15

Fig. 2a illustrates one embodiment of a packet 200 that may be used in the embodiment of Fig. 1. Packet 200 may include a header 202 and a payload 204. Header 202 may include sequencing and network routing information, indicators of the source and destination devices, an address corresponding to a beginning data block, and an indicator of the number of data blocks involved in the request that corresponds to the packet. Payload 204 may include one or more data blocks being transferred by packet 200. The number of data blocks in payload 204 may be variable or fixed according to a given communications protocol. It can be noted that payload 204 may be empty (or absent) for certain types of packets. Header 202 and/or payload 204 may include other information as required by a protocol.

25

In one embodiment, network 110 may implement a bandwidth-reserving wormhole routing model. In this model, header 202 may travel ahead of its

corresponding, variable-size payload 204 in network 110. As the header routes from switching device to switching device in network 110, it may reserve bandwidth (and/or other resources) for its payload 204 at each subsequent switching device. Once the resource are reserved in a switching device, the switching device can assure that payload
5 204 will route successfully to the next switching device. If a switching device along the way cannot provide the required resources for payload 204, then that switching device may be configured to drop payload 204 and generate an intention packet as discussed above. In this embodiment, the switching device may generate the intention packet using header 202 or part of header 202.

10

Fig. 2b illustrates one embodiment of an intention packet 222. Intention packet 222 may include intention information 226, severed flag 228 and diagnostic information 230. In one embodiment, intention information 226 may include the information of header 202 of the dropped packet. In other embodiments, intention information 226 may
15 include only selected or modified portions of header 202 of the dropped packet. Severed flag 228 may include an indicator that the intention packet corresponds to a dropped packet. Diagnostic information 230 may include a switch indicator which may identify the switching device where the original packet was dropped. Diagnostic information 230 may include congestion statistics which may include congestion information of the
20 switching device where the original packet was dropped. The intention packet may be generated by the device using software, hardware or a combination thereof.

Once a switching device generates intention packet 222, the switching device may convey intention packet 222 to the destination device through network 110. Since
25 intention packet 222 may comprise a small packet, it may be assured (or nearly assured) of routing through the subsequent switching devices in network 110 to reach its destination device if congestion was the adverse condition. The switching devices may be configured to insure that they always have the required bandwidth or resources to route

packets up to a maximum size. The intention packets may be below this maximum size such that the switching devices may always route them.

5 A device may react to receiving an intention packet based on the information in the intention packet as well as its own characteristics. The information in the intention packet may indicate whether the dropped packet was a request or a response to a request. If the dropped packet was a request, then the device that receives the intention packet, i.e. the receiving device, may send a signal to the requesting device that its request has been dropped. This signal may be a NACK packet conveyed from the receiving device so that
10 the requesting device may resend the dropped packet. The receiving device may also perform an operation that corresponds to the request in anticipation of receiving the resent packet as discussed below in more detail.

If the dropped packet was a response to a request, then the requesting device will
15 receive the intention packet. The requesting device may reschedule the request in response to receiving the intention packet and may convey a subsequent packet that includes the request to the receiving device. The requesting device may be configured to optimize the resending of the request by rescheduling the request for a more appropriate time or by merging the request with a subsequent request.

20

In one particular embodiment of Fig. 1, device 100 may comprise a host interface device that is coupled to a computer bus or network (not shown) and device 120 may comprise a disk drive. In this embodiment, the host interface device may perform read and write accesses to the disk drive on behalf of devices in the computer network. The
25 cases of a write request and a read request with the host interface and disk drive will now be discussed to illustrate the use of intention packets in a congested network.

00320-182960

In one embodiment, when the disk drive receives a non-severed request, i.e. a request that is not a resent request, the disk drive may compare the non-severed request to the list of severed requests. If the non-severed request is independent of each of the severed requests, e.g. attempts to read or write a different region of data than each of the severed requests, then the disk drive may complete an operation corresponding to the non-severed request. If the non-severed request is not independent of one or more of the severed requests, then the disk drive may be configured not to complete the operation corresponding to the non-severed request and may convey a NACK to the host interface that sent the non-severed request. For example, each request may have a sequence number. If a non-severed read request for a location arrives at the disk drive and the severed requests list indicates that a severed write request with an earlier sequence number is pending for the same location, then the non-severed read request will be delayed until the severed write request is resent and completed. However, if the non-severed read request is independent from the requests in the severed requests list, then it will be allowed to complete, even if it arrives out of sequence.

When the host interface receives a NACK from the disk drive, either indicating that its request has been dropped or that its request is not independent of another request that has been severed, the host interface may reschedule the request and may convey a new packet corresponding to the request to the disk drive. The host interface may optimize the rescheduled request in a number of ways. For example, the host interface may reschedule the request at a less congested time. In addition, the host interface may merge the request with a subsequent request while ensuring consistency.

To perform a read request, a host interface may convey a packet corresponding to the read request to the disk drive through network 110. The disk drive may receive the read request packet and may read the data block or blocks specified by the request. The

disk drive may then convey the data block or blocks in a response packet to the host interface. If each switching device encountered by the response packet has sufficient resources to transmit the packet, then the response packet may be routed to the host interface to complete the read request. If the resources required by the response packet are not available at a switching device in network 110, then that switching device may drop the response packet and generate an intention packet. The intention packet may include information from the header of the response packet as well as a severed flag and diagnostic information, as shown in Fig. 2b. The switching device may then convey the intention packet to the host interface though network 110. Again, the intention packet may be assured (or nearly assured) of avoiding further congestion to the host interface by virtue of its relatively small size.

Note that in some cases, an adverse condition or combination of adverse conditions may be severe enough so that the intention packet cannot be routed or transmitted to its destination or cannot be routed or transmitted to its destination in a reasonable time. Such cases may be handled for the intention packet by conventional mechanisms, such as a time-out mechanism.

When the host interface receives an intention packet, the host interface may reschedule the read request and may convey a new packet corresponding to the read request to the disk drive. The host interface may optimize the rescheduled request in a number of ways. For example, the host interface may reschedule the request after the disk drive has spun around to the location of the request or it may retarget the read request to a less congested mirror copy. In addition, the host interface may merge the request with a subsequent request while ensuring consistency.

Turning now to Fig. 3, a flow chart illustrating a method for implementing a routing scheme employing intention packets in a switching device is shown. Variations

on the method are possible and contemplated. In Fig. 3, a first packet that includes a header and a payload may be received as indicated in block 302. In block 304, a determination may be made if an adverse routing condition is present. For example, a determination may be made as to whether required routing resources (such as bandwidth and buffering space) are available to transmit the payload to a second device. If no adverse condition is detected, then the header may be conveyed to the second device as indicated in block 306 and the payload may be conveyed to the second device as indicated in block 308. It can be noted that the header and payload may be conveyed to the second device at substantially the same time or at separate times depending on a routing model used by a system.

If an adverse condition is detected, such as the required routing resources are not available to transmit the payload to the second device, then the payload may be dropped as indicated in block 310. An intention packet may be generated as indicated in block 312. The intention packet may include all or part of the header of the original packet along with other information as discussed above. The intention packet may be conveyed to the second device as indicated in block 314.

Turning now to Fig. 4, a flow chart illustrating a method for handling packets at a first exemplary device is shown. The first exemplary device may comprise a disk drive or another type of device. Variations on the method are possible and contemplated. In Fig. 4, a packet may be received at a receiving device as indicated in block 402. In block 404, a determination may be made as to whether the packet is an intention packet. In one embodiment, an intention packet may include a flag to indicate that it is an intention packet. If the packet is an intention packet, then intention packet information may be recorded as indicated in block 406. In addition, an operation corresponding to the intention packet may be begun as indicated in block 408, such as prefetching data blocks or flushing a write buffer. A NACK may be conveyed to the sending device as indicated

in block 410. Note that the elements of the method indicated at 408 and 410 do not necessarily occur serially. Element 410 may be performed before 408, after 408 or concurrently with 408.

5 If the packet is not an intention packet, then a determination may be made as to whether an operation corresponding to the packet is independent of an outstanding operation as indicated in block 412. The outstanding operation may correspond to other intention packets that were received at the receiving device prior to the packet. If the operation corresponding to the packet is independent of the outstanding operations, or
10 alternatively, if there are no outstanding operations at the receiving device, then the operation corresponding to the packet may be begun as indicated in block 414. If the operation corresponding to the packet is not independent of the outstanding operations, then a NACK may be conveyed to the sending device as indicated in block 416. In addition, the operation corresponding to the packet may not be performed.

15 Turning now to Fig. 5, a flow chart illustrating a method for handling packets at a second exemplary device is shown. The second exemplary device may comprise a host interface or another type of device. Variations on the method are possible and contemplated. In Fig. 5, a first packet corresponding to a request from a requesting
20 device may be conveyed as indicated in block 502. A second packet corresponding to a response to the request may be received at the requesting device as indicated in block 504. In block 506, a determination may be made as to whether the second packet is an intention packet. In addition, a determination may be made as to whether the second packet is a NACK as indicated in block 508. If the second packet is either an intention
25 packet or a NACK, then the request may be rescheduled as indicated in block 510 and a third packet corresponding to the request may be conveyed from the requesting device as indicated in block 512.

It can be noted that the requesting device may receive an intention packet when the response to its request was dropped in route to the requesting device. The response, for example, may have included data corresponding to a read request from the requesting device. It can be further noted that the requesting device may receive a NACK when its request was not independent of an outstanding request at a receiving device.

Turning now to Fig. 6, a block diagram of one embodiment of a system is shown. Fig. 6 depicts a plurality of application servers 600a through 600(n) (collectively, application servers 600), file system server 610, and disk array 630 coupled to a network 606. Disk array 630 includes controller 632a and a plurality of disk drives 636a through 636(n) (collectively, disk drives 636) coupled to network 634. In addition, controller 632b is coupled to network 606 and disk array 630.

The embodiment of Fig. 6 may be configured to implement the routing scheme discussed above with respect to Figs. 1-5. In particular, controller 632a and controller 632b may each operate as a host interface to network 606 and may perform accesses to disk drives 636 on behalf of application servers 600 and file system server 610. Controller 632a and controller 632b may convey packets to and receive packets from the disk drives 636 through network 634. In response to congestion or other adverse conditions at a switching device in network 634, a packet may be dropped and an intention packet may be generated as discussed above. The intention packet mechanism may also be applied to network 606.

Turning now to Fig. 7, a block diagram of one embodiment of a plurality of nodes connected in by a network is shown. Any number of nodes 710 may be included in the network. In the embodiment of Fig. 7, each node 710 may include a switching device 720. Nodes 710 may also include one or more devices such as a disk drive 730 or local host (e.g. computing) system 740. In other embodiments, each node 710 may include

other types or numbers of devices and the nodes 710 may not be identical. The above-described routing mechanism employing intention packets may be applied to the system of Fig. 7. For example, a device at one node may initiate a request to a device at another node, such as from one host system to another or from a host system to a disk drive. A request may be routed from one node to another through several of the switching devices. If routing problems are encountered, the above-described intention packet routing mechanism may be employed.

Various embodiments may further include receiving or storing instructions and/or data implemented in accordance with the foregoing description upon a carrier medium. Suitable carrier media may include storage media or memory media such as magnetic or optical media, e.g., disk or CD-ROM, as well as transmission media or signals such as electrical, electromagnetic, or digital signals, conveyed via a communication medium such as network 110 and/or a wireless link.

Although the embodiments above have been described in considerable detail, other versions are possible. Numerous variations and modifications will become apparent to those skilled in the art once the above disclosure is fully appreciated. It is intended that the following claims be interpreted to embrace all such variations and modifications.

I CLAIM:

1. An apparatus comprising:

5 a first device;

a second device;

a switching device coupled to said first device and said second device;

10

wherein said switching device is configured to receive a first packet from said first device, wherein said switching device is further configured to detect an adverse transmission condition, wherein in response to an adverse transmission condition, said switching device is configured to drop at least
15 a portion of said first packet, generate a second packet, and convey said second packet to said second device, wherein said second packet includes information about said first packet.

15

2. The apparatus of claim 1, wherein said first packet includes a header and a
20 payload, and wherein said switching device is configured to drop said payload and generate said second packet from at least a portion of said header.

20

3. The apparatus of claim 1, wherein said second packet includes a flag indicating that at least a portion of said first packet has been dropped.

25

4. The apparatus of claim 1, wherein said second device is configured to convey a third packet to said first device in response to receiving said second packet.

11. The apparatus of claim 4, wherein said first packet corresponds to a response to a request, wherein said second device is configured to reschedule said request in response to receiving said second packet, and wherein said third packet comprises the rescheduled request.

5

12. The apparatus of claim 1, wherein said first device comprises a host interface, and wherein said second device comprises a disk drive.

13. The apparatus of claim 1, wherein said adverse transmission condition comprises congestion above a threshold at said switching device is configured to convey said first packet.

10

14. The apparatus of claim 1, wherein said adverse transmission condition comprises corruption of said first packet.

15

15. The apparatus of claim 1, wherein said adverse transmission condition comprises a fault condition preventing routing of said first packet along an intended route.

16. A method comprising:

20

receiving a first packet from a first device;

determining if an adverse transmission condition is present; and

25

in response to determining that said adverse transmission condition is present:

dropping at least part of said first packet;

generating a second packet, wherein said second packet includes
information about said first packet; and

conveying said second packet to a second device.

5

17. The method of claim 16, wherein said first packet includes a header and a
payload, wherein said dropping comprises dropping said payload, and wherein said
generating comprises converting said header into a severed header.

10 18. The method of claim 16, wherein said generating comprises generating a flag that
indicates that at least part of said first packet has been dropped.

19. The method of claim 16, further comprising conveying a third packet to said first
device in response to receiving said second packet.

15

20. The method of claim 19, further comprising rescheduling a request in response to
receiving said third packet, wherein said first packet corresponds to said request.

21. The method of claim 20, wherein said third packet comprises a negative
20 acknowledgement.

22. The method of claim 19, further comprising performing a first operation
corresponding to a request indicated by said second packet, wherein said first packet
corresponds to said request.

25

23. The method of claim 22, further comprising:

receiving a fourth packet subsequent to receiving said second packet;

determining whether said second operation is independent of said request; and

performing a second operation corresponding to said fourth packet in response to
5 determining that said second operation is independent of said request.

24. The method of claim 23, further comprising:

conveying a fifth packet corresponding to said fourth packet in response to
10 determining that said second operation is not independent of said request,
wherein said fifth packet comprises a negative acknowledgement.

25. The method of claim 22, further comprising:

15 receiving a fourth packet subsequent to receiving said second packet; and

performing a second operation corresponding to said fourth packet to complete
said request.

20 26. The method of claim 19, further comprising:

rescheduling a request in response to receiving said second packet, wherein said
first packet corresponds to a response to said request, and wherein said
third packet corresponds to said request.

25

27. The method of claim 16, wherein said first device comprises a host interface, and
wherein said second device comprises a disk drive.

28. The method of claim 16, wherein said adverse transmission condition comprises congestion above a threshold.

29. The method of claim 16, wherein said adverse transmission condition comprises corruption of said first packet.

30. The method of claim 16, wherein said adverse transmission condition comprises a fault condition preventing routing of said first packet along an intended route.

31. A device comprising:

a first port configured to receive a first packet;

a second port;

a switching device coupled to said first port and said second port, wherein said switching device is configured to receive said first packet from a first device at said first port, wherein said switching device is further configured to detect an adverse transmission condition, wherein in response to an adverse transmission condition, said switching device is configured to drop at least a portion of said first packet, generate a second packet, and convey said second packet from said second port to a second device, wherein said second packet includes information about said first packet.

32. The device of claim 31, wherein said second packet includes at least a portion of a header from said first packet.

[illegible]

[illegible]

5
10
15
20

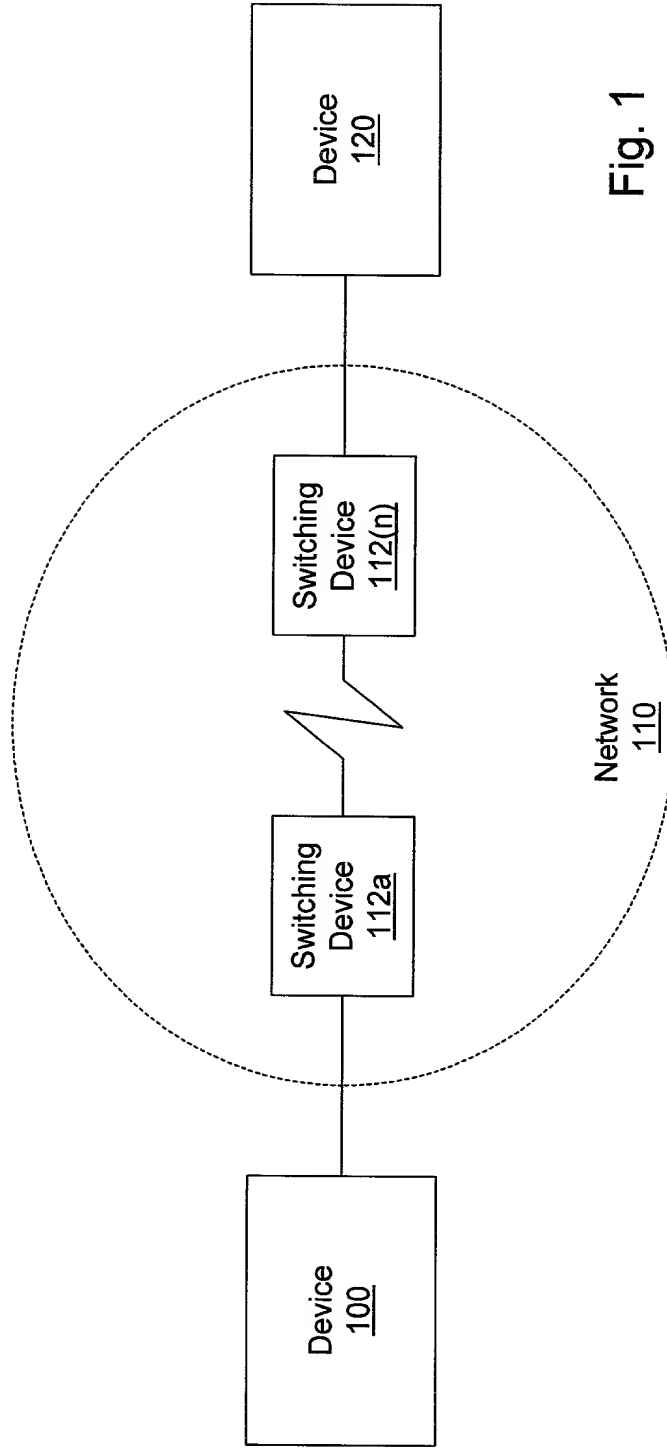


Fig. 1

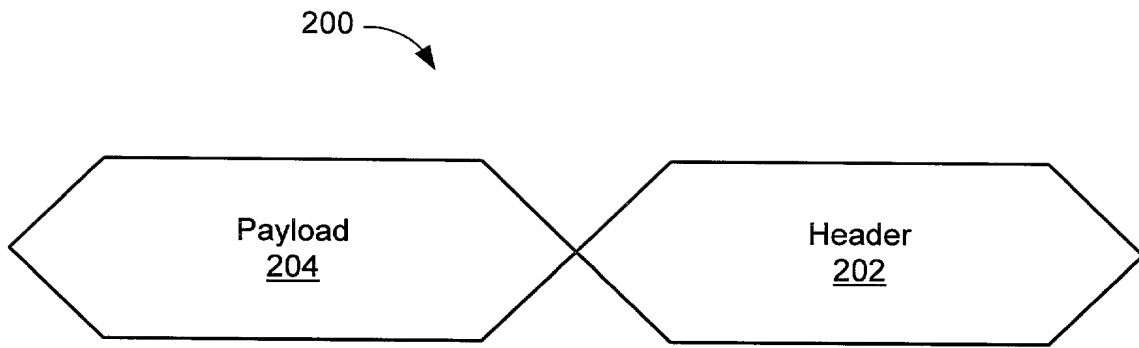


Fig. 2a

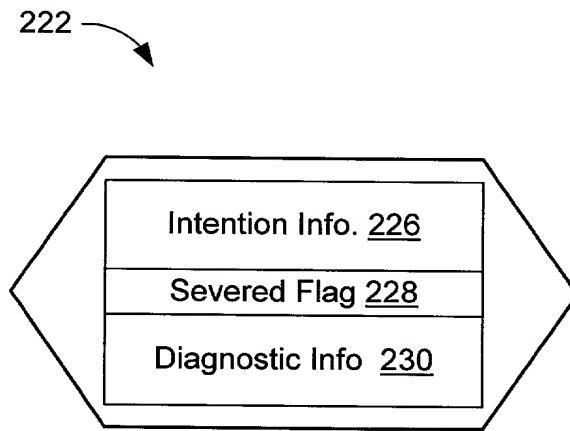


Fig. 2b

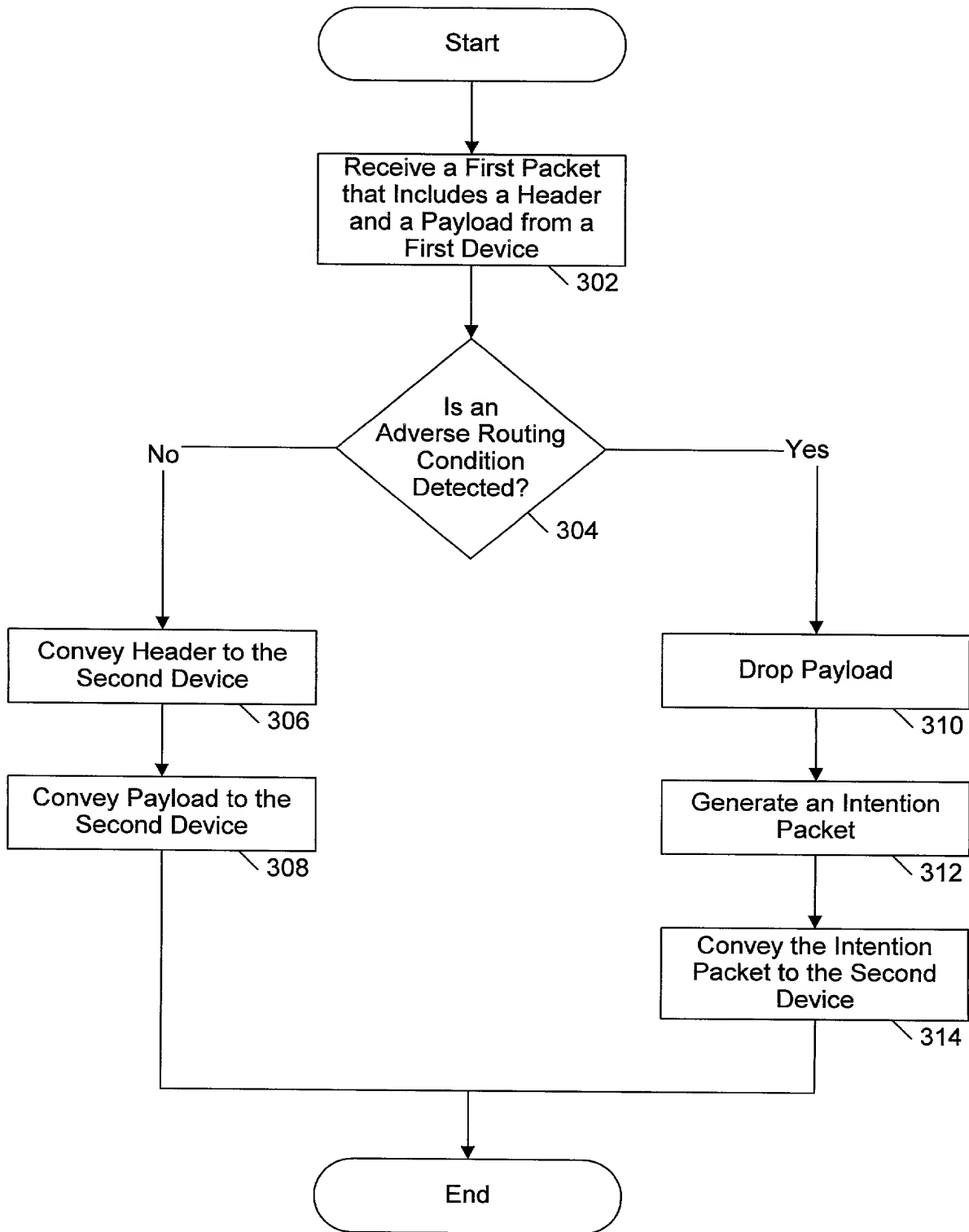


Fig. 3

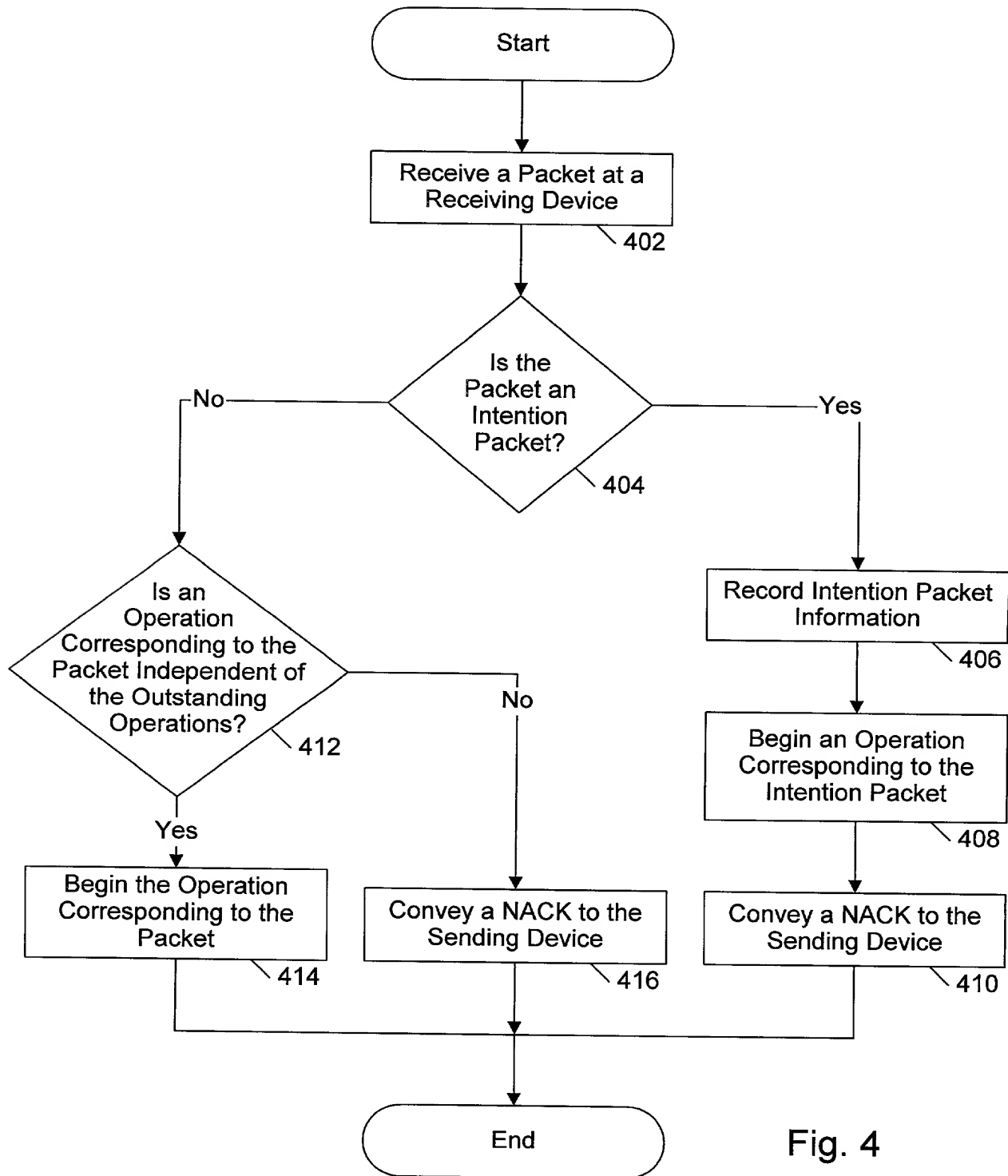


Fig. 4

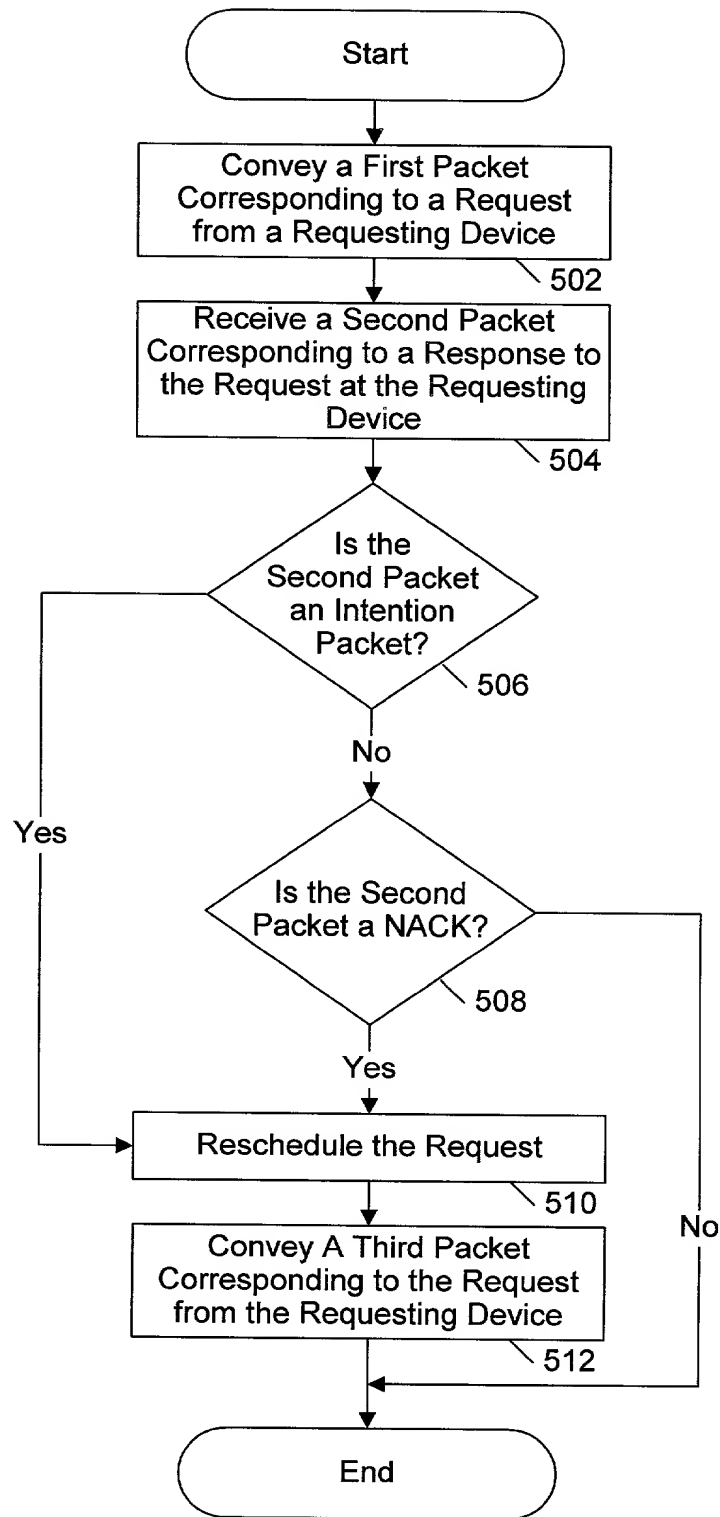


Fig. 5

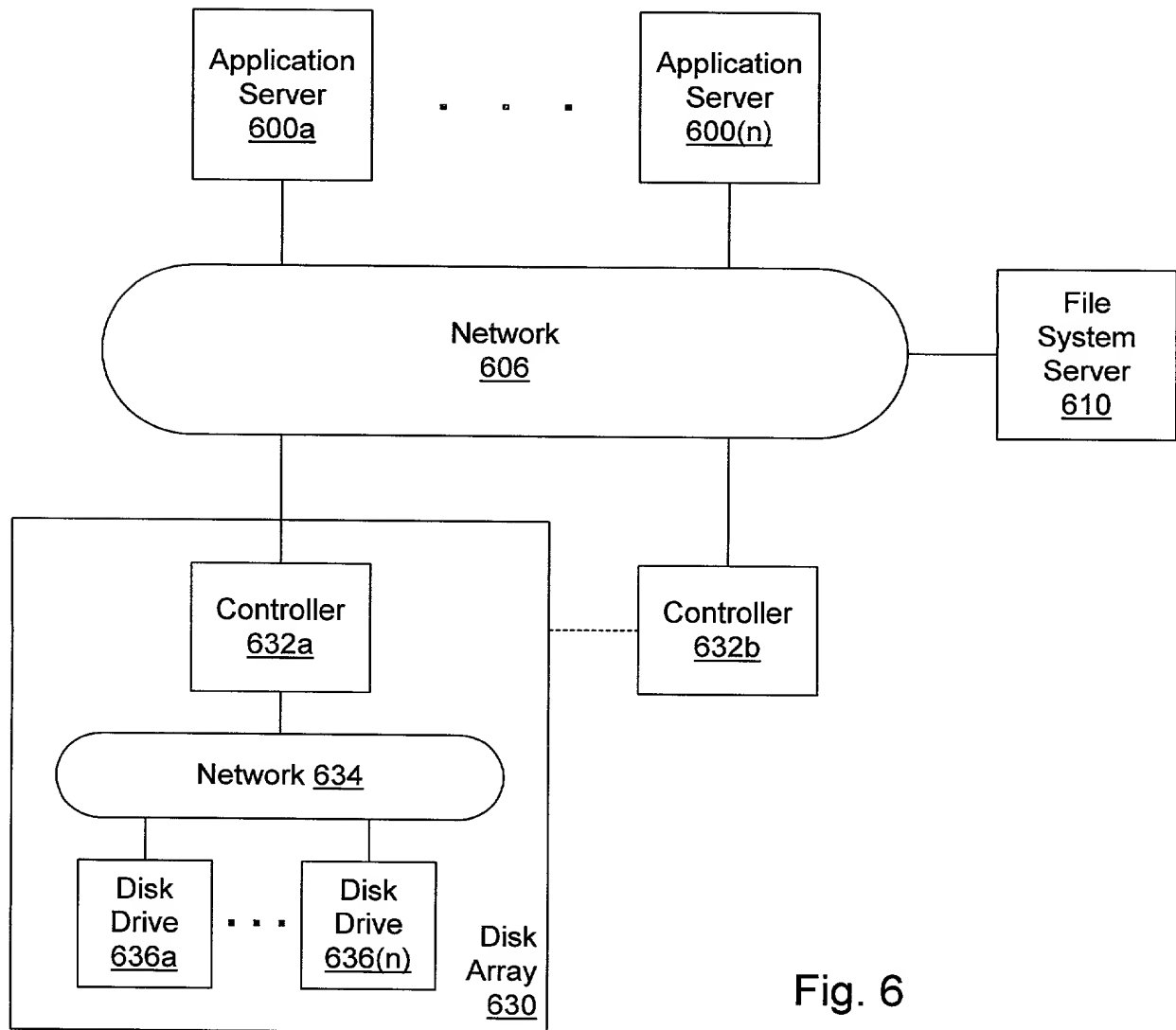


Fig. 6

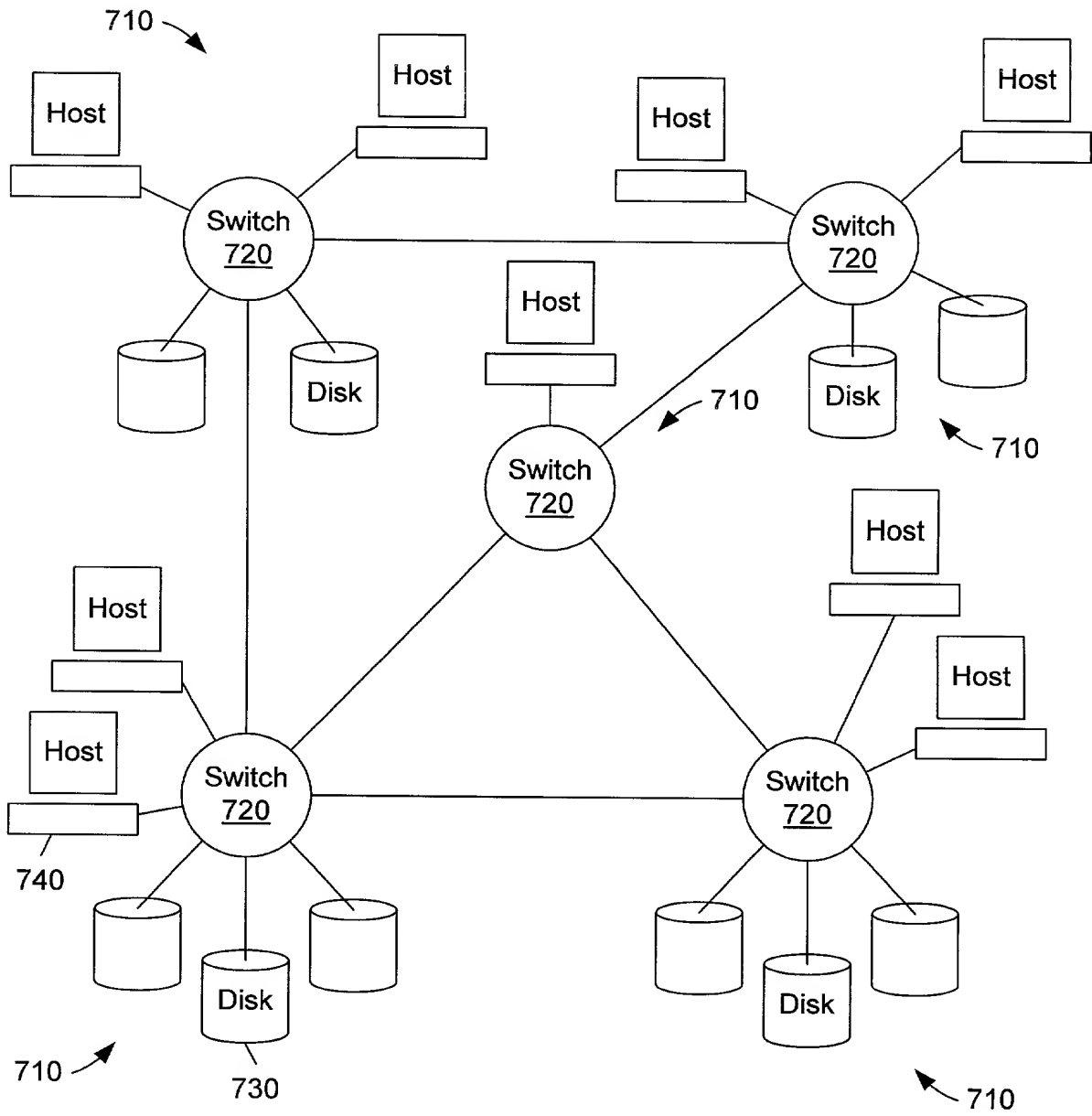


Fig. 7

DECLARATION AND POWER OF ATTORNEY

As a below named inventor, I hereby declare that:

My residence, post office address, and citizenship are as stated below next to my name.

I believe I am the original, first and sole inventor (if only one name is listed below) or an original, first and joint inventor (if plural names are listed below) of the subject matter which is claimed and for which a patent is sought on the invention entitled "**SYSTEM AND METHOD FOR IMPLEMENTING A ROUTING SCHEME USING INTENTION PACKETS IN A COMPUTER NETWORK**" the specification of which:

- ☒ is attached hereto.
☐ was filed on _____ as Application Serial No. _____
 and was amended on _____ (if applicable).

I hereby state that I have reviewed and understand the contents of the above-identified specification, including the claims, as amended by any amendment referred to above.

I acknowledge the duty to disclose to the Patent and Trademark Office all information known to me to be material to patentability of the subject matter claimed in this application, as "materiality" is defined in 37 C.F.R. § 1.56.

I hereby claim foreign priority benefits under 35 U.S.C. § 119(a)-(d) or § 365(b) of any foreign application(s) for patent or inventor's certificate listed below, or under § 365(a) of any PCT international application listed below designating least one country other than the United States of America, and have identified below any foreign application for patent or inventor's certificate, or of any PCT international application, having a filing date before that of the application on which priority is claimed.

<u>Prior Foreign Application No.</u>	<u>Country</u>	<u>Filing Date</u> (mm/dd/yy)	<u>Priority</u> <u>Claimed</u>	<u>Cert. copy</u> <u>Attached</u>
N/A				

I hereby claim the benefit under 35 U.S.C. § 119(e) of any United States provisional application(s) listed below.

<u>Provisional Application No.</u>	<u>Filing Date</u> (mm/dd/yy)
N/A	

I hereby claim the benefit under 35 U.S.C. § 120 of any United States application(s) listed below, or under § 365(c) of any PCT international application listed below designating the United States of America, and, insofar as the subject matter of each of the claims of this application is not disclosed in the prior United States or PCT international application in the manner provided by the first paragraph of 35 U.S.C. § 112, I acknowledge the duty to disclose all information known to me to be material to the patentability of the subject matter claimed in this application, as "materiality" is defined in 37 C.F.R. § 1.56, which became available between the filing date of the prior application and the national or PCT international filing date of this application.

<u>Parent Application No.</u>	<u>Filing Date</u> (mm/dd/yy)	<u>Parent Patent No. (if applicable) or Status</u>
N/A		

I hereby revoke any previous Powers of Attorney and appoint

Kenneth Olsen	Reg. No. 26,493
Robert S. Hauser	Reg. No. 37,847
Alexander E. Silverman	Reg. No. 37,940
Christine S. Lam	Reg. No. 37,489
Timothy J. Crean	Reg. No. 37,116
Joseph T. FitzGerald	Reg. No. 33,881
Anirma Rakshpal Gupta	Reg. No. 38,275
Sean P. Lewis	Reg. No. 42,798
Michael J. Schallop	Reg. No. 44,319
Bernice B. Chen	Reg. No. 42,403
Kenta Suzue	Reg. No. 45,145
Noreen A. Krall	Reg. No. 39,734
Richard J. Lutton, Jr.	Reg. No. 39,756
Monica D. Lee	Reg. No. 40,696
Marc D. Foodman	Reg. No. 34,110
Naren Chaganti	Reg. No. 44,602

each of said attorneys being employed by Sun Microsystems; and

Dan R. Christen	Reg. No. 39,943
Gentry E. Crook	Reg. No. 44,633
Kevin L. Daffer	Reg. No. 34,146
Mark R. DeLuca	Reg. No. 44,649
Jeffrey C. Hood	Reg. No. 35,198
Robert C. Jahnke	Reg. No. 44,800
B. Noël Kivlin	Reg. No. 33,929
Robert C. Kowert	Reg. No. 39,255
Lawrence J. Merkel	Reg. No. 41,191
Eric B. Meyertons	Reg. No. 34,876
Louise K. Miller	Reg. No. 36,609
David W. Quimby	Reg. No. 39,338
David A. Rose	Reg. No. 26,223

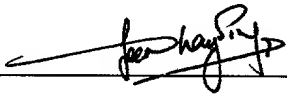
each of said attorneys or agents being a member or an associate of the firm of Conley, Rose & Tayon, P.C., as attorney or agent for so long as they remain with such company or firm, with full power of substitution and revocation, to prosecute the application, to make alterations and amendments therein, to transact all business in the Patent and Trademark Office in connection therewith, and to receive the Letters Patent.

Please direct all communications to:

Robert C. Kowert
Conley, Rose & Tayon, P.C.
P.O. Box 398
Austin, Texas 78767-0398
Phone: (512) 476-1400

I hereby declare that all statements made herein of my own knowledge are true and that all statements made herein on information and belief are believed to be true; and further that these statements were made with the knowledge that willful false statements and the like so made are punishable by fine or imprisonment, or both, under 18 U.S.C. § 1001 and that such willful false statements may jeopardize the validity of the application or any patent issued thereon.

Inventor's Full Name: Whay S. Lee

Inventor's Signature:  Date: July 18, 2000

City and State (or Foreign Country) of Residence: Newark, CA Citizenship: Malaysia

Post Office and Residence Address: 37171 Sycamore St., Apt. 1438, Newark, CA 94560
(Include number, street name, city, state and zip code)

0052381-03000